LIVING IN HOKKAIDO

Cynthia Jung

May 2021

IBM APPLIED DATA SCIENCE CAPSTONE PROJECT

I. Introduction

Background

When people usually talk about visiting Japan, you will most likely hear them talking about Tokyo. However, some people who want to visit Japan might want to visit somewhere calmer. Japan is made up of 5 different islands and is divided into 47 prefectures. From parks to restaurants, there are many exciting places to explore throughout Japan.

Business Problem

This project will be focusing on the second largest island of Japan that also comprises the largest prefecture of Japan, Hokkaido. If a person wanted to move here, checking out locations is the best way to ensure you will enjoy where you are going. Some things that you might look for when deciding if it is the right place to visit, could be something like, if there are places to visit like museums or parks. Or maybe you have always wanted to visit a hot spring. Checking the location ahead of time will help you determine if the things that interest you are in this city and if not are easy to travel to from wherever you decide to go.

II. Data

Data Description

The objective of this capstone project is to analyze and recommend where someone would like to visit or possibly even move to. Where someone might stay or live, and what kind of culture they would be surrounded by. The ease of access to various places such as markets, airports, train stations, restaurants, entertainment and other daily needs. In order to help someone

do so, you will first need to know where they want to go and from there narrow down their search by finding out what is around that area. Once you know where you want to go, using Foursquare API we can search around the area for things to do.

Resources

Data Link: https://en.wikipedia.org/wiki/Subprefectures_of_Hokkaido

This link contains a list of subprefectures in Hokkaido. Using this data along with data obtained from Foursquare API and Geocoder we will be able to determine venue data about specific locations throughout Hokkaido.

Using the Data

To solve the problem, we will need to collect the following data:

1. Determine and find the latitude and longitude of the capital city of Hokkaido.

2. Get information about the subprefectures in Hokkaido.

3. Select a subprefecture and get the name of its capital.

4. Collect venue data for the subprefecture.

5. Find the top 10 venues.

6. Find out how many unique types of the venues there are.

After collecting all the necessary data, we can use k-means clustering and various map visualizations to make it easier to understand where everything is located and how close they are to each other. This way, you can quickly see where you would want to stay if you wanted to be in the center of all the activity or somewhere a little farther away for some peace and quiet.

III. Methodology

The first step we need to do in order to begin is get information about Hokkaido. First, Hokkaido is part of Japan. Hokkaido is the second largest island of Japan. Next, we need to understand how Hokkaido is divided. Fortunately for us, there is a friendly chart available on a Wikipedia page (https://en.wikipedia.org/wiki/Subprefectures_of_Hokkaido). For us to utilize the information provided by this webpage, we will need to do some web scraping. To this, we will be using various libraries and packages on Python such as requests and BeautifulSoup in order to extract the table's data.

Once we have extracted that information, we then need to find the geographical coordinates in the form of latitude and longitude. This information is extremely important as we will need it in order to be able to use Foursquare API. To acquire the geographical coordinates that we will be using for the Foursquare API, we will be using the package called Geocoder. This package will allow us to convert an address into latitude and longitude.

After gathering this data, we will need to rearrange it into a data frame. This is done using a library in Python known as pandas. Once the data is organized into a data frame, then we can use it to visualize the information on a map. Python makes it easy for use to visualize data using what is called a Folium package. With this package, we can create interactive leaflet maps.

Next, we will need to use Foursquare API to get information about the venues around a specific city in Hokkaido. For this project, we will be focusing on a city called Asahikawa. The first thing we must do in order to use Foursquare API is register for a Foursquare

Developer Account. Once you have registered, you will be given a client ID and a client secret key. This are necessary to access the information through the Foursquare API. Now to use Foursquare API to acquire the top 10 venues within a designated radius we need to use the geographical coordinates we previously acquired. Foursquare will return this venue data as a JSON file. From that file, we can obtain venue names, categories, latitude and longitude. With this data we can analyze each city in Asahikawa and examine how many different types of venues there are in the area. As we analyze this data, we are preparing it to be used for clustering.

Finally, our last step is clustering all the data using k-means clustering. We will be clustering the data into 5 clusters based on the frequency of a venue's occurrence. K-means clustering is an algorithm which identifies k number of centroids and then allocates every data point given to the nearest centroid, thus creating clusters. The goal is to make sure that the data is accurately distributed. The results will allow use to easily see where there are going to be a lot of stuff going on compared to somewhere else which maybe a bit quieter and more spread out from various activities.

IV. Results

The results from k-means clustering shows us that we can cluster the venues in Asahikawa into 5 clusters based on the frequency of occurrence.
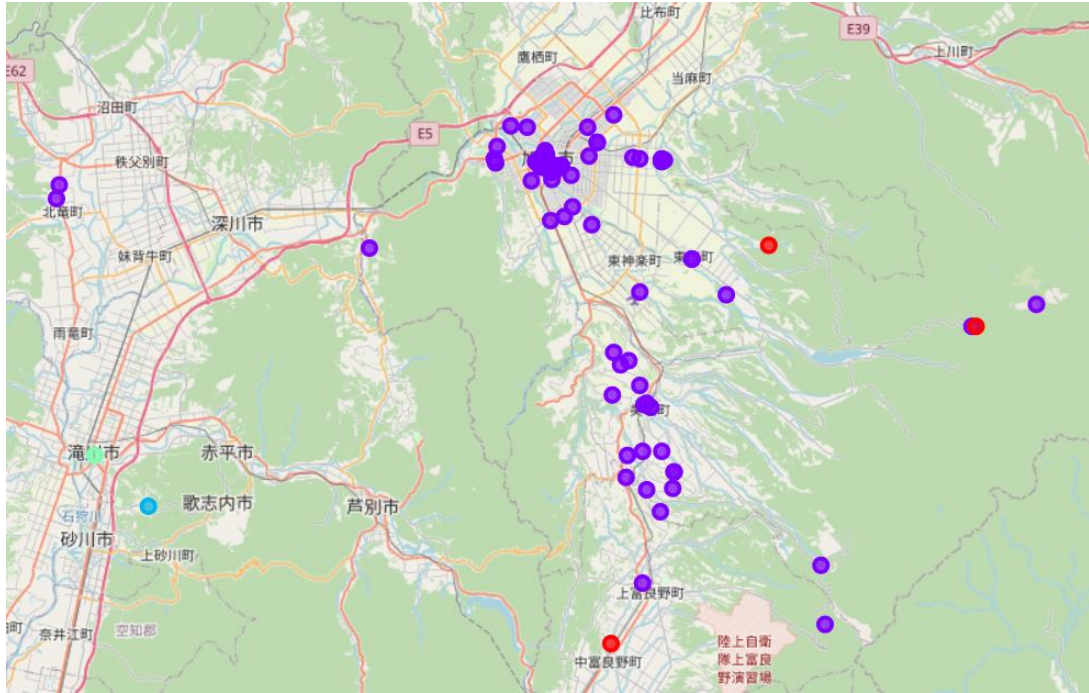
Based on the clusters, we can determine where there are many venues in a specific location compared to others. This is extremely helpful if a person who plans to live in Hokkaido wants somewhere quiet to live or someone who wants to make sure there are a lot of place to going shopping for daily goods nearby.

V. Discussion

Based on observations made from the map in the Results section, majority of the venues are located towards the center of a subprefecture. In this case we are looking at the subprefecture Kamikawa, whose capital is Asahikawa. As you start moving away from the city, you will notice that the variety of venues nearby decrease immensely. However, various forms of transportation such as the train or undersea railway, makes getting around Hokkaido and other parts of Japan more easily accessible.

Therefore, if you prefer a higher paced environment, an area with a higher concentration of venues would be highly recommend for someone who is outgoing and does not mind the noise and chaos that comes with a densely populated area. However, if you want a change of pace from the hustle and bustle of city life, it would be more beneficial if you moved to an area that isn't in the center of many venues. Instead, moving somewhere a little a farther away from the capital but still near a train station, might be just what you are looking for. Somewhere that is quiet but not completely cut off from the venues only found in major populated areas.

VI.   Conclusion

There are many components to completing this project. In order to find a solution to the business problem, we have gone through an exceptionally long process. From acquiring and preparing the data, to preforming clustering and analyzing results. There were many steps that needed to be taken in order to determine the best location for a potential new homeowner. Therefore, based on the findings of this project living at the center of a heavily dense cluster is best for those who are outgoing. While the lightly dense clusters are more ideal for people who prefer to stay indoors.